TITLE

NOVEL HUMAN α1 CHAIN COLLAGEN

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates to a novel human collagen protein and a polynucleotide sequence, which encodes the novel human collagen protein. More particularly, it relates to polynucleotides encoding human α1 chain collagen and derivatives thereof.

Description of the Related Art

Collagens are structure proteins that participate in the assembly of various kinds of polymers in the extracellular matrix. Collagen polypeptides contain one or more blocks of (Gly-x-y) repeat, in which x represents any amino acid residue, and y frequently represents prolyl or hydroxyprolyl residues. The presence of such sequence repeats allows groups of three collagen polypeptides to fold into triple-helical domains, which are rigid and inextensible.

So far, 20 distinct types of homo- and heterotrimeric molecules, encoded by more than 30 genes, have been identified in vertebrates. These proteins exhibit considerable diversity size, sequences, tissue distribution, molecular composition, and each plays a different structure role in connective tissue.

Within the superfamily of collagens, two categories are classified. The fibrillar collagens include types I, II, III, V, and XI collagen. The triple-helical domains of the

proteins polymerize in a staggered fashion to form fibrils. Members of other collagens do not by themselves form cross-striated fibrils, but may be associated with fibrils (FACIT or fibril associated collagens with interrupted triple helices), including types IX, XII, XIV, XVI, and XIX collagen. The structure of these molecules comprises two or more relatively short triple-helical (COL) domains connected and flanked by non-triple-helical (NC) sequences. Type IX collagen is the best-characterized molecule in the members. Studies of transgenic mice with mutations in type IX collagen have been proposed that it acts as molecular bridges between cartilage collagen fibrils and other matrix components, perhaps proteoglycans. The COL domains and the central NC domains of this molecule interact with type II collagen through covalent cross-links to form fibrils. The amino-terminal NC domain has a potential of interacting with other extracellular components. Also, in vitro studies have demonstrated that the N-terminal non-triple-helical domains of type XII and XIV collagen promote contraction of collagen gels. However, the detailed interactions of the bridging hypothesis are not clear.

Collagens are typical mosaic proteins containing a number of shuffled domains. These domains have been classified by sequence similarity in order to characterize their structural and functional relationships to other proteins. This analysis provides an overview of homologies of collagen domains. It also reveals two new relationships: (i) a module common to type V, IX, XI, and XII collagens was found to be homologous to the heparin binding domain of thrombospondin; (ii) the modular architecture of a human

type VII collagen fragment was identified. Its N-terminal globular domain contains fibronectin type III repeats located adjacent to a von Willebrand factor type A module. The proposed structural similarities point to analogous subfunctions of the respective domains in otherwise distinct proteins.

Thrombospondin is one of a class of adhesively homotrimeric glycoproteins that mediate cell-to-cell and cell-to-matrix interactions. It is expressed in extracellular matrix, and may have autocrine growth regulatory properties involved in platelet aggregation, embryogenesis, morphogenesis, cell adhesion molecule, major activator of TGFβ1. The von Willebrand factor A (vWF) like domain is the prototype for a protein superfamily and it is found in various proteins including plasma complement factors, integrands, collagens, and other extracellular proteins. Proteins that incorporate vWF domains participate in numerous biological events, such as cell adhesion, migration, homing, pattern formation, and signal transduction.

Collagens are important bio-medical building blocks with the functions of tissue growth, anaplasty, dressing for burn, and wound healing, etc., and the requirements thereof expand largely. Therefore, there is still a need to develop a novel collagen and the derivatives thereof having more therapeutic value and diversity for the various applications.

The present inventors have successfully cloned a novel human α1 chain collagen gene by way of known human expression sequence tag (EST) in combination with bio-informatics and molecular cloning techniques. After

comparison of the inventive collagen with the existent 20 collagens, the highest sequence homology is less than 30%, indicating the collagen of the invention is a novel form.

Blood vessels are tubes of endothelial cells surrounded by layers of smooth muscle cells and connective tissue proteins. During development this complex structure forms as a result of biochemical signals between endothelial cells and smooth muscle cells. Sometimes this biochemical communication fails and abnormal blood vessels form. By analyzing gene mutations causing such vascular abnormalities, it can be learned about the signals necessary for normal blood vessel development. In addition, identification of genes responsible for inherited vascular malformations provides a basis for development of rational therapies in the clinical treatment of vascular disorders.

SUMMARY OF THE INVENTION

It is therefore the primary object of the present invention to provide an isolated nucleic acid (*hCOLA1*) and the degenerate sequences thereof, which encodes human α1 chain collagen protein, comprising the nucleotide sequence set forth in SEQ ID NO. 5. The present invention also provides the expression profile of the isolated collagen gene and the exact tissue and cellular localization of this collagen protein. Moreover, the present invention provides nucleotide fragments derived from SEQ ID NO. 5 as a nucleic acid probe or primer.

In one preferred embodiment, the present invention provides a novel human α1 chain collagen protein encoded by

the nucleic acid mentioned above, which has the amino acid sequence set forth in SEQ ID NO. 1.

Another aspect of the present invention provides a recombinant vector comprising the nucleic acid mentioned above and a regulatory sequence.

Still another aspect of the present invention provides a method for producing human α1 chain collagen protein, comprising the steps of: (a) transforming or transfecting a host cell with the recombinant vector described above; (b) culturing said transformed or transfected cell under the conditions sufficient for expression of the human α1 chain collagen protein; and (c) recovering and purifying the human α1 chain collagen protein.

Yet still another aspect of the present invention provides a diagnostic kit for detecting the disease related to the mutation of SEQ ID NO. 5 in a mammal or human, comprising the nucleic acid probe or primer described above.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will be more fully understood and further advantages will become apparent when reference is made to the following description of the invention and the accompanying drawings in which:

FIG. 1 is a diagram showing the PCR cloning for the human α1 chain collagen cDNA of the invention, wherein lane 1 is the molecular weight markers; lane 2 is the negative control in which the human cDNAs are devoid; and lane 3 is the PCR product containing the cDNA coding for the human α1 chain collagen of the invention.

FIG. 2 is a diagram showing the construct of the recombinant vector Bluescript KS(+)/*E. coli* (*hCOLA1*) of the present invention.

FIG. 3 is a diagram showing the complete nucleotide sequence (SEQ ID NO. 5) and the corresponding amino acid sequence (SEQ ID NO. 1) of the human α1 chain collagen of the invention.

FIG. 4 is a diagram showing the hydropathy profile of the deduced amino acid sequence of SL. Kyte-Doolittle hydrophobicity profile of the human α1 chain collagen plotted with a 11-residue window.

FIG. 5 is a schematic diagram showing domain structure of the human α1 chain collagen protein of the invention.

FIG. 6(A-C) is a diagram showing the amino acid sequence comparison of the human α1 chain collagen protein of the invention with type IX (Col9a1) and type XIX (Col19a1) collagens wherein the black box refers to the identical amino acid residue, and symbol "-" refers to gap insertion for the optimal alignment. The lower panel of FIG. 6 also shows the evolution tree among these collagens.

FIG. 7(A) is a Northern blot containing 2 μg of poly(A)$^+$ RNA from indicated tissues hybridized with human α1 chain collagen cDNA-specific probe; and human β-actin-specific probe as an internal control. FIG. 7(B) is a Northern blot containing 2 μg of poly(A)$^+$ RNA from indicated cardiovascular tissues hybridized with human α1 chain collagen cDNA-specific probe; and human glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) probe as an internal control.

FIG. 8 is a quantitative RT-PCR of the expression of human α1 chain collagen from human fetal and adult tissues.

Human glyceraldehyde 3-phosphate dehydrogenase was used as internal control.

FIG. 9 is an in situ hybridization analyses of expression of the human α1 chain collagen mRNA expression. Cardiovascular sections and cells were hybridized with digoxigenin labeled antisense riboprobes for human α1 chain collagen. (A) Longitudinal section, artery; (B) longitudinal section, ventricle; and (C) aortic smooth muscle cells. Control hybridizations labeled with sense probes did not produce signals (data not shown). *Bar*, 10 μm.

FIG. 10 is a diagram showing the expression of the human α1 chain collagen protein in *E. coli*. FIG. 10(A) shows SDS-PAGE analysis, wherein the numbers indicated are molecular weight standards; lane 1 is the non-induced cell lysate; lane 2 is the cell lysate induced by IPTG for 2 hours; and lane 3 is the cell lysate induced for 3 hours. In FIG. 10(B), lane 1 shows the human α1 chain collagen protein purified by Ni-column and stained with Coomassie brilliant blue; and lanes 2 and 3 are western blot detected by anti-histidine antibody, wherein lane 2 is the non-induced cell lysate and lane 3 is the cell lysate induced by IPTG for 2 hours without purification.

FIG. 11 is a diagram showing RT-PCR of the recombinant expression of human α1 chain collagen in COS7 cells, wherein "-" refers to negative control; "+" is the RT-PCR products from transformants containing human α1 chain collagen gene; and numbers indicated are molecular weight standards.

DETAILED DESCRIPTION OF THE INVENTION

The present invention screened the most conserved regions of the known collagen nucleic acid sequences from human expressed sequence tag (EST) library. A novel bio-molecule with the highest homology of the primary amino acid sequence was then found by introducing the sequence of that region into EST library. A full-length cDNA sequence of the novel bio-molecule was then obtained and determined by the technologies of bio-informatics and molecular cloning.

At the beginning, a 57-bp fragment of the conserved region was aligned in the human EST library to obtain a fragment with about 300 bp in length. The fragment was then introduced into Genbank Blast for searching human non-redundant genes to obtain a fragment with about 146 kb in length containing exons and introns. Possible open reading frames were analyzed and the relative oligonucleotide probes were thus designed to clone the novel full-length human α1 chain collagen. The method of cloning will be further described in the following examples.

The nucleic acid sequence of the full-length human α1 chain collagen (*hCOLA1*) gene and the deduced amino acid sequence thereof are shown in FIG. 3 (SEQ ID NO. 5 and SEQ ID NO. 1, respectively). The novel human α1 chain collagen gene comprises 2,865 bp, which encodes 954 amino acids with about 99,000 Da in molecular weight, and is located at the p11.2-12.3 on human chromosome XI.

The above isolated nucleic acid (*hCOLA1* gene) comprises at least the nucleotide sequence set forth in SEQ ID NO. 5 (including DNA and RNA sequences) or the complementary sequences thereto, and the genomic DNA sequence. Those

skilled in this art will be aware that the nucleotide sequences can be modified in accordance with any method known in the art, and are also within the scope of the invention. For example, degenerate codons can be used to replace the relative positions but the gene encodes the same amino acid sequence. Further, additional codons can be inserted into the sequence or added either at the 3'- or 5'-end, but the activity of the protein is not affected or slightly affected. Accordingly, the complementary sequences and degenerate sequences of SEQ ID NO. 5, and various modified variants are included in the present invention. See, for example, Sambrook, et al., Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York, 1989.

In accordance with the present invention, the human α1 chain collagen encoded by the isolated *hCOLA1* gene comprises three domains (as shown in FIG. 5), including (i) von Willebrand factor A domain (having amino acid sequence set forth in SEQ ID NO. 2); (ii) thrombospondin N-terminal-like domain (having amino acid sequence set forth in SEQ ID NO. 3); and (iii) collagenous domain (having amino acid sequence set forth in SEQ ID NO. 4). To analyze the primary sequence of the protein and to compare with other known collagens, the human α1 chain collagen of the present invention belongs to FACIT family (fibril associated collagens with interrupted triple helices). From the domains described above, it is inferred that the physiological functions of the human α1 chain collagen of the invention may be involved in platelet aggregation, cell adhesion, and the activation of transformation growth factor. In addition,

the N-terminal of the collagen protein further includes a signal peptide with 22 amino acids in length. It is inferred that the human α1 chain collagen of the invention is located in the extracellular matrix.

The present invention also provides a recombinant vector comprising the nucleic acid cloned and isolated above, and optionally a regulatory sequence, such as replication region, selection marker (e.g. antibiotic resistance marker), and promoter, etc. The promoter used herein can be an eukaryotic cell promoter or a prokaryotic cell promoter so that the recombinant vector can be expressed in a suitable host, for example, eukaryotic cells such as mammalian cells or yeast, or prokaryotic cells such as *Escherichia coli*.

In one preferred embodiment, the present invention provides a recombinant vector in which the *hCOLA1* gene is cloned into Bluescript KS (+) vector (Strategene). The recombinant vector is deposited at the Culture Collection and Research Center (Hsinchu, Taiwan) on November 14, 2000, and assigned accession number CCRC 940331.

One can produce the novel human α1 chain collagen protein which have the amino acid sequence set forth in SEQ ID NO. 1 using the isolated nucleic acid described above by any suitable method in any suitable expression system known in this art. Therefore, the method for producing human α1 chain collagen protein is also within the scope of the present invention.

One preferred expression system for the recombinant production of the collagen of the invention is in transgenic non-human animals, wherein the desired collagen may be recovered from the milk of the transgenic animal. Such a

system is constructed by operably linking the DNA sequence encoding the collagen of the invention to a promoter and other required or optional regulatory sequences capable of effecting expression in mammary gland. Likewise, required or optional post-translational enzymes may be produced simultaneously in the target cells, employing suitable expression system operable in the targeted milk protein producing mammary gland cells.

In one preferred embodiment of the present invention, the nucleic acid of SEQ ID NO. 5 is subcloned into an expression vector to obtain another recombinant vector. A suitable host cell (for example, eukaryotic or prokaryotic cell) is then transformed or transfected with the recombinant vector. The transformed or transfected cells are then cultured under the conditions sufficient for expression of human α1 chain collagen protein. Finally, the expressed proteins are recovered and purified. Those skilled in this art will appreciate that the recovering and purifying method is not limited, for example, by various chromatographies. Preferably, the human α1 chain collagen is expressed using histidine tag fusion protein technique, and the recovering and purifying method is performed by affinity column.

As used herein, the term "transformation" or "transfection" includes a variety of techniques for introducing an exogenous nucleic acid into a cell (for example, eukaryotic or prokaryotic), including calcium phosphate or calcium chloride precipitation, microinjection, DEAE-dextrin-mediated transfection, lipofection, or electroporation.

Electroporation is carried out at approximate voltage and capacitance (and corresponding time constant) to result in enter of the DNA construct(s) into the host cells. Electroporation can be carried out over a wide range of voltages (e.g. 50 to 2,000 volts) and corresponding capacitance. Total DNA of approximately 0.1 to 500 μg is generally used.

Methods such as calcium phosphate precipitation and colubrine precipitation, liposome fusion and receptor-mediated gene delivery can also be used to transfect cells.

The genetic engineering methods mentioned above such as DNA modification, cloning, construction, and isolation of the recombinant vector, protein expression, and purification can be accomplished by those skilled in this art, and which can be seen in, for example, Ausubel F. M., et al., Current Protocols in Molecular Biology, New York, 1992; Sambrook, et al., *supra*; or Davis, L. G., Methods in Molecular Biology, Elsevier, Amsterdam, NL, 1986.

In one aspect of the present invention, the isolated nucleic acid further includes the fragments derived from SEQ ID NO. 5 or the complementary sequences thereto to be as a nucleic acid probe or primer for detection. Those skilled in the art will be aware that the length of the nucleic acid fragment is not limited. For example, as a nucleic acid probe, the fragment preferably comprises at least 500 contiguous nucleotides in length derived from SEQ ID NO. 5 or more, while as a nucleic acid primer, the fragment preferably comprises at least 20 contiguous nucleotides in length derived from SEQ ID NO. 5 or more. The selection of the length of fragment is dependent upon the conditions of

detection method as described below. For example, the temperature and ionic strength used in hybridization, or the temperature used in polymerase chain reaction (PCR). Generally, the length of the nucleic acid probe is in proportion to the specificity of the detection result. Accordingly, the nucleic acid probe preferably comprises at least 500 contiguous nucleotides in length derived from SEQ ID NO. 5, and more preferably comprises the full-length nucleic acid of SEQ ID NO. 5. In addition, the length of the nucleic acid primer is in proportion to the specificity of the detection result. Accordingly, the nucleic acid primer preferably comprises at least 20 contiguous nucleotides in length derived from SEQ ID NO. 5, and more preferably comprises 20-25 contiguous nucleotides, thereby increasing the specificity of the detection result.

The human α1 chain collagen polynucleotide of the present invention may be used for diagnostic and/ or therapeutic purposes. For diagnostic uses, the polynucleotide of the invention may be used to detect the human α1 chain collagen gene expression or aberrant α1 chain collagen gene expression in disease states, e.g., rheumatoid arthritis, osteoarthritis, reactive arthritis, autoimmune bearing disease, cartilage inflammation due to bacterial or viral infections (e.g. Lyme's disease), parasitic disease, bursitis, corneal diseases, ankylosing spondylitis (fusion of the spine), and cardiovascular disease.

In the present invention the inventor suggested the novel collagen is derived from blood vessels, and maybe relates to cardiovascular disease. By analyzing gene

mutations causing cardiovascular abnormalities, it can provide a basis for development of rational therapies in the clinical treatment of cardiovascular disorders.

The kit of the present invention used for detecting such diseases comprises a probe or primer described above. Methods for detecting the expression of *hCOLA1* gene by using the nucleic acid probe include, but are not limited to, Northern analysis, Southern analysis, *in situ* hybridization, and bio-chip/microarray, etc., which are well known in the art. Those skilled in the art will appreciate that methods using the complementary properties between two nucleic acid molecules are within the scope of the present invention. In addition, methods for detecting the expression of *hCOLA1* gene by using the nucleic acid primer include, but are not limited to, reverse transcriptase polymerase chain reaction (RT-PCR), 5'-Rapid Amplification of cDNA End (5'-RACE), and 3'-RACE, etc. Those skilled in the art will appreciate that methods using the at least one primer in combination with PCR are also within the scope of the present invention.

The human α1 chain collagen gene and/or protein of the present invention may be useful in the treatment of various abnormal conditions. By introducing gene sequences into cells, gene therapy can be used to treat conditions in which the cells underexpress normal α1 chain collagen or express abnormal/inactive α1 chain collagen. In some instance, the polynucleotide sequence encoding the human α1 chain collagen of the invention is intended to replace or act in the place of a functionally deficient endogenous gene. Alternatively, abnormal conditions characterized by overproliferation can be treated using the antisense of the

human α1 chain collagen coding sequence of the invention. Recombinant gene therapy vectors, such as viral vectors, may be engineered to express the human α1 chain collagen of the invention. Thus recombinant gene therapy vectors may be used therapeutically for treatment of diseases resulting from aberrant expression or activity of the human α1 chain collagen of the invention.

Without intending to limit it in any manner, the present invention will be further illustrated by the following examples.

EXAMPLE

Example 1. cDNA cloning of *hCOLA1*

A Clontech SMART RACE cDNA Amplification kit was used to clone *hCOLA1* cDNA. Sequence specific primers used for the following RACE reactions were either deduced from the previously published partial human genomic clone 682J15 (Genbank Accession No. AL034452) or the cloned *hCOLA1* cDNA fragment. Initially, first strand cDNA was synthesized from 1 μg of total RNA pool (Clontech) using Superscript II reverse transcriptase with a specific primer 5'-GGTTCACCTTTGCTTCCCTTAG-3', deduced from the clone 682J15. The reaction was following to the manufacture's protocol. The above reverse transcription reaction mixture was used for 5'RACE reaction with a sequence specific primer (5'-TTGGCCCATTAATCCTCGGTTTC-3'), corresponding to nucleotides 1823-1845 of the *hCOLA1* cDNA and the universal primer provided by the kit. All assays were performed in a 50-μl reaction volume using the GeneAmp PCR system 9600 (Perkin-Elmer Cetus).

To obtain the entire coding region of *hCOLA1* gene, first strand cDNA was synthesized from 1 μg of total RNA pool (Clontech) using Superscript II reverse transcriptase with an oligo dT primer. After reverse transcription, 1 μl of the reaction mixture was used for PCR amplification with a upstream primer (5'-ATTCCTGGGCCACCTGGTCCGATA-3'), corresponding to the most 5' candidate initiator methionine of the clone 708F5 (Genbank Accession No. AL031782) and a downstream primer (5'-CTAATAGTTTGGTCCTTTTCT-3'), corresponding to the 3' stop codon of the clone 682J15. A single band with a molecular size of 2.9 kilo bases was obtained (Figure 1). The band was excised from gel and cloned into the BlueScript II KS(+) vector (Strategene). The recombinant vector was deposited at the Culture Collection and Research Center (Hsinchu, Taiwan) and assigned accession number CCRC 940331. After nucleotide sequence analysis, the PCR product was found to contain the entire open reading frame of *hCOLA1*.

Example 2. Nucleotide sequencing

Nucleotide sequencing was carried out with the Sanger dideoxynucleotide chain termination method (Sambrook, et al., 1989). The sequence samples were prepared using the Ampli Taq cycle sequencing kit (Perkin-Elmer, Inc.) following the manufacturer's protocol. The samples were applied to a 377 automatic sequencer (Perkin-Elmer, Inc.). All reported sequences were confirmed by sequencing of both sense and antisense strands. The full-length nucleotide sequence (SEQ ID NO. 5) and the deduced amino acid sequence (SEQ ID NO. 1)

of the human α1 chain collagen of the invention is shown in FIG. 3.

Example 3. Northern blot analysis

The human multiple tissue and the cardiovascular Northern blots, containing 2 μg of poly(A)$^+$ RNA from indicated tissues, were obtained from Clontech (catalog number 7780-1 and 7791-1, respectively). The blot was hybridized with a randomly primed ^{32}P-labeled probe corresponding to nucleotides 1236-1863 of the *hCOLA1* open reading frame at 60°C in ExpressHyb solution for one hour and washed with 2× SSC/0.1% SDS two times for 15 min each at 60°C. Then the blot was washed with 0.2× SSC/0.1% SDS three times for 15 min each at 60°C. Human *β actin* or *GAPDH* probe was used as a control for the amount of RNA in each lane. As shown in Fig. 7A, a transcript of approximately 4.3 kb is observed, in agreement with the size of the cloned cDNA. The expression of *hCOLA1* collagen is mostly confined to placenta and heart tissues, with lower levels in skeletal muscle, small intestine, liver and lung. Another transcript of approximately 2.4 kb was detected to be hybridized with the probe in heart tissue. It probably is a splicing variant of the *hCOLA1* gene. We further examined the expression pattern of *hCOLA1* in human cardiovascular tissues containing fetal heart and adult heart tissues, together with the aortic and cardiac tissues by Northern blot analysis. Surprisingly, the *hCOLA1* transcripts were only present in fetal heart and aortic tissues (Fig. 7B). Moreover, the 2.4 kb short transcript was only present in the fetal heart. Another 7.3 kb band was detected in both

tissues. We do not know if this is an additional splicing variant of the *hCOLA1* gene. No hybridization signal was detected in adult heart and cardiac tissues. Although the result showing the absence of *hCOLA1* transcript in adult heart is inconsistent with the data of Northern blot analysis in Fig. 7A, the *hCOLA1* mRNA level in fetal heart is 22-fold in excess of the adult heart based on the quantitative RT-PCR results in Fig. 8 (see below). The presence of the *hCOLA1* transcripts in aorta suggests that this novel collagen is derived from blood vessels.

Example 4. Quantitative RT-PCR

Five micrograms of total RNAs from a variety of human fetal and adult tissues obtained from Clontech (catalog number K4005-1) were used for reverse transcription reactions with oligo (dT) primers. After reverse transcription reactions, the relative quantity of endogenous *GAPDH* mRNA in each tissue sample was determined with CYBR Green fluorescence dye (Molecular Probes) using Real-time PCR analysis (LightCycler, Roche Molecular Biochemicals). The resulting *GAPDH* mRNA value in each tissue sample was used to normalize the sample for differences in the amount of total RNA added to each PCR reaction. Each of the normalized tissue samples was then split to perform the target α1(XXI) collagen and control *GAPDH* amplifications by Real-time PCR analysis. The relative quantity of α1(XXI) collagen cDNA in each reaction was determined in the exponential phase to ensure that the amount of product amplified reflects the quantity of starting mRNA. Primers used for PCR amplifications are as follows: *GAPDH* (5'-

TGAAGGTCGGAGTCAACGGATTTGGT-3' and 5'-CATGTGGGCCATGAGGTCCACCAC-3'; 983-bp fragment); *hCOLA1* collagen (5'-TTCCTGGAAACCGAGGATTAATG-3' and 5'-AGTCCACGATCACCCTTGTCAC-3'; 1546-bp fragment). Meanwhile, samples at a PCR cycle in the linear range of amplification (30 cycles for *hCOLA1*; 20 cycles for *GAPDH*) were electrophoresed on 1.5% agarose and stained with ethidium bromide for visualization. As shown in Fig.8, when normalized to the *GAPDH* values, the relative amounts of *hCOLA1* transcripts were 2.7, 22 and 30 times more in fetal brain, heart and liver than in the adult counterparts, respectively. The results indicate that *hCOLA1* expression is developmentally regulated and suggest a role for α1(XXI) collagen in developmental processes in multiple tissues. Comparison of the *hCOLA1* expression in different adult tissues reveals that high levels of *hCOLA1* expression were detected in trachea, testis, uterus, and placenta, with modest levels of expression in brain, lung, colon, prostate, spinal cord, and salivary gland. The *hCOLA1* collagen mRNA expression was very low or undetectable in adult heart, liver, kidney, bone marrow, spleen, thymus, skeletal muscle, and adrenal gland.

Example 5. In situ hybridization analysis

In situ hybridization was performed on 5-μm human cardiovascular tissue sections (Novagen, catalog number 70316-3). An antisense or sense RNA probe labeled with digoxigenin-UTP (DIG-UTP) encompassing the region corresponding to nucleotides 1236-1547 of the *hCOLA1* open reading frame was obtained by *in vitro* transcription

(Boehringer RNA labeling kit). Sections were dewaxed by washing three times for 5 min in xylene. After dewaxing, sections were rehydrated to PBS through an ethanol series, washed three times in PBS, and then incubated for 15 min in a proteinase K solution (10 µg/ml in PBS). Proteinase K activity was stopped by washing twice in PBS and sections were refixed at RT for 30 min in 4% paraformaldehyde, 0.2% glutaraldehyde in PBS. After fixation, sections were washed twice in PBS then incubated for 1 h at 50°C in pre-hybridization mix (50% foramide, 5X SSC, 50 µg/ml yeast tRNA, 0.1% SDS and 50 µg/ml heparin). Hybridization mix containing probe was replaced and incubated at 50°C for overnight. After hybridization, sections were washed twice for 30 min at 50°C in solution I (50% foramide, 5 X SSC, and 0.1%SDS) and twice for 30 min at 50°C in solution II (50% foramide, 2 X SSC, and 0.1% SDS). Sections were washed three times at RT in MAB (100 mM maleic acid, 150 mM NaCl, pH 7.5) and then blocked for 2 h at RT with 2% blocking reagent (Boehringer) in MAB. Sections were incubated for 2 h at RT with 2% blocking reagent in MAB containing 1: 2000 dilution of anti-DIG antibody. The sections were washed 4 times for 15 min at RT in MAB-Tween (0.1% Tween-20), washed three times for 5 min in AP buffer (0.1 M Tris-HCl, pH 9.0, 50 mM $MgCl_2$, 0.1 M NaCl, and 0.1% Tween). Color was developed by incubating the sections in NBT/BCIP. After developing, sections were washed in PBS and counterstained with nuclear fast red and then mounted with Histomount Mounting Solution (Zymed). Cells grown on coverslips for *in situ* hybridization analysis were performed according to the previously published protocol.

Example 6. Expression of *hCOLA1* in *Escherichia coli* and purification

The entire coding region of the *hCOLA1* cDNA was amplified by PCR with primers 5'-ATGGCTCACTATATTACATTTCTC-3', corresponding to the 5' cDNA region and 5'-TTAGTGATGGTGATGGTGATGCTCATAGTTTGGTCCTTTTCTG-3', corresponding to the 3' region including 6 histidine residues right before the stop codon. The amplified DNA construct was gel purified and sub-cloned into the expression vector pET 15b (Novagen) in which the Nco I site was digested and blunted with Klenow fragment. The recombinant protein was obtained by expressing the constructs in *E. coli* strain BL21 (DE3). The transformed *E. coli* was cultured in LB medium containing 100 μg/ml of ampicillin at 37 °C to reach an optical density of 0.7 at 600 nm, followed by induction with IPTG at a final concentration of 1 mM and kept culturing for an additional 2 or 3 hours. The cell lysate with total proteins was analyzed by SDS-PAGE. The result is shown in FIG. 10(A).

One liter of the IPTG induced *E. coli* cells was cultured for 2 hours and then centrifuged at 5000 xg for 30 min. The cell pellet was washed with PBS and centrifuged again. The cell pellet was then re-suspended in 20 ml of PBS containing 1 mM of PMSF. The cell suspension was subjected to ultrasonication to break the cell walls. The cell lysate was then centrifuged at 30,000 xg for 40 min. The supernatant was applied to a Ni-agarose column (5 ml in bed volume) that has been equilibrated with 50 mM of Tris-HCl buffer, pH 8.0 at a flow rate 0.5 ml/min. The column

was washed with the same buffer containing 40 mM of imidazole. The recombinant *hCOLA1* was eluted with the same buffer containing 0.25 M imidazole. The eluate was quantified and analyzed by SDS-PAGE, followed by staining with Coomassie brilliant blue. A protein band with 98 kDa in molecular weight was observed on the gel (FIG. 10(B), lane 1). In addition, the proteins without purification were blotted to a PVDF membrane. An antibody to histidine tag (Clontech) was used to detect the recombinant protein. The result of Western blot is shown in FIG. 10(B), lanes 2 and 3, in which the band indicated at 98 kDa corresponds to the human α1 chain collagen protein of the invention.

Example 7. Expression of *hCOLA1* in eukaryotic cell

The *hCOLA1* cDNA containing entire open reading frame prepared by Example 4 was gel purified and sub-cloned into the expression vector pcDNA 3.1 containing CMV promoter (Invitrogen) in which the Pme I site was digested and blunted with Klenow fragment. The mammalian cells COS7 were transfected with the expression vector via Superfect (Qiagen), and cultured in DMEM supplemented with 10% FBS (Life Technologies) for 48 hours. About 10^6 cells were used for the extraction of total RNA. The reverse transcription was performed with oligo dT primer using 0.2 μg RNA as template. After reaction, PCR was carried out with primers T7 and BGHrev on the pcDNA3.1 vector using 0.5 μl solution. The result is shown in FIG. 11, indicating that the vector is expressed in the transfected mammalian cells.

Referring to FIG. 4, the first 22 amino acid residues indicated by a solid bar encode a putative signal peptide characterized by secreted proteins. It is inferred that the human α1 chain collagen protein of the invention is located in the extracellular matrix.

The amino acid sequence of the human α1 chain collagen protein of the invention is compared with those of other 20 known collagens, particularly type IX and type XIX, the most similar in structures. The amino acid sequence identity of collagens between type IX and *hCOLA1* of the invention is 24%, while that between type XIX and *hCOLA1* of the invention is 27%, indicating that *hCOLA1* of the invention is a novel form of collagen (FIG. 6).

While the invention has been particularly shown and described with the reference to the preferred embodiment thereof, it will be understood by those skilled in the art that various changes in form and details may be made without departing from the spirit and scope of the invention.